MISCOVERGENCE IN THE LANCZOS ALGORITHM

(Abbreviated title: Misconvergence)

Beresford Parlett[*]

Department of Mathematics
Computer Science Division of the
Department of Electrical Engineering and Computer Science
University of California
Berkeley, California 94720

In memory of James H. Wilkinson

ABSTRACT.   The Lanczos algorithm generates Ritz values in order to approximate eigenvalues. If some eigenvalues are clustered then a Ritz value may hover at a wrong value for a good number of steps. We study this phenomenon and focus on the point of discovery, the first step at which it is certain that there is a hidden eigenvalue in the vicinity of stabilized Ritz values. Both before and after this point the Ritz value behavior is routine — but for different eigenvalue configurations.

The "effective spread" at step j is an interval guaranteed to contain all unknown eigenvalues. The notion of "Ritz intervals" leads to a computable counterpart to the exact theory.

AMS Category: Primary 65F15

1. INTRODUCTION

When the Lanczos algorithm is applied to a symmetric n×n matrix A it produces, at step j, an unreduced tridiagonal matrix T_j whose eigenvalues

$$\theta_1^{(j)} < \theta_2^{(j)} < \ldots < \theta_j^{(j)} \quad ,$$

will be called the jth set of Ritz values. The final set of Ritz values are the eigenvalues of A,

$$\lambda_1 = \theta_1^{(n)}, \ldots, \lambda_n = \theta_n^{(n)} \quad .$$

We may assume, without loss of generality, that the $\{\lambda_i\}$ are all distinct from each other. The characteristic polynomial of T_j is

$$\chi_j(\xi) = \prod_{i=1}^{j} (\xi - \theta_i^{(j)}) \quad .$$

One attraction of the Lanczos algorithm is that in most cases for small values of j, but j ≥ max{10, √n}, *some* extreme Ritz values from the jth set are very good approximations to *some* extreme eigenvalues of A.

Error bounds on $|\theta_i^{(j)} - \lambda_i|$ were given in (Kaniel 1966; Paige 1972; Saad 1980). Here we concentrate on a related question concerning occasional eccentric behavior of the Ritz values. When a cluster of (unknown) eigenvalues has a spread that is 1% or less of neighboring eigenvalue gaps then a Ritz value can hover at the mean of the cluster for 8 or more successive values of j before making an abrupt shift to one of the eigenvalues. This phenomenon is called *misconvergence* (see Parlett et al., 1982). It was studied very carefully in (van der Sluis and van der Vorst 1987) and we shall refer frequently to this paper.

We offer a simple description of this odd Ritz value behavior that complements the detailed error bounds in (S. and V. 1987). We find it necessary to look beyond the Ritz values near the cluster for a simple

explanation and are lead to the *effective spread* which gives the allow-
able range for the unknown eigenvalues at each step. This concept is
useful in describing the actual convergence rate of the conjugate
gradient algorithm.

After some preliminaries the paper is organized as follows. The
example in (S. and V. 1987) and a snapshot of the paper's many results
is followed by our formal description in Section 5. Then, by using
Paige's persistance theorem, we can offer a constructive theory that
avoids reference to unknown quantities. Finally we present a theorem
which shows that the special value of j at which the presence of a
hidden eigenvalue is detected (the point of discovery) can be character-
ized by the occurrence of a local minimum in the error bound of the
misleading stagnant Ritz value.

The figures and charts are worth a thousand words.

2. ORTHOGONAL POLYNOMIALS

By Cauchy's interlace theorem (Parlett 1980, Chap. 10 or S. and V. 1987)
each set of Ritz values interlaces the next set. Moreover, the charac-
teristic polynomials are orthogonal with respect to a special discrete
inner product function

$$\langle \phi, \psi \rangle := \sum_{k=1}^{n} w_i \phi(\lambda_i) \psi(\lambda_i) \quad .$$

Each positive weight $\{w_i\}$ is the squared cosine of the angle between the
i^{th} eigenvector of A and the starting vector of the Lanczos algorithm.
Thus the whole set of Ritz values is determined by the inner product and
this point of view is the fundamental one. It follows from orthogonality
that each χ_j is minimal (Szegö 1939):

$$\|\chi_j\| = \min_{\phi} \|\phi\|$$

over all monic polynomials ϕ of degree not exceeding j. Here

$$\|\phi\| := \langle \phi, \phi \rangle^{\frac{1}{2}} \ .$$

The drawback of this approach is that in the Lanczos context $\langle \ , \ \rangle$ is unknown and the χ_j are built up one at a time.

We shall employ both viewpoints in this paper.

3. THE DUTCH EXAMPLE

The following values are used in (S. and V. 1987).

$$\lambda_1 = .0340 \ , \quad \lambda_2 = .0341 \ , \quad \lambda_3 = .0820 \ ,$$

$$\lambda_4 = .127 \ , \quad \lambda_5 = .155 \ , \quad \lambda_6 = .190 \ ,$$

$$\lambda_7, \ldots \lambda_{900} \text{ uniformly spaced in } [.2, \ 1.2].$$

$$w_i = 1/900, \quad i = 1, \ldots, 900 \ .$$

It should be emphasized that λ_1 and λ_2 are not to be considered pathologically close. The gap poses no difficulties to current software when invoked using numbers with at least 14 decimal digits. The question here is how many more steps does the Lanczos algorithm require to find λ_1 and λ_2 than to find λ_3 ?

Figure 1 shows those Ritz values that are near λ_1 and λ_2 for $j = 7, \ldots, 30$. For $j = 20, \ldots, 25$

$$.03404952 < \theta_1^{(j)} < .03404999 \ ,$$

so the changes during this stagnation phase are smaller than 1% of the actual error, $\theta_1^{(j)} - \lambda_1$.

In (S. and V. 1987) the behavior of $\theta_1^{(j)}$ is divided into three phases:

i) normal (rapid) convergence of θ_1 to the mean $(\lambda_1 + \lambda_2)/2$, $j = 1, \ldots, 18$.

ii) stagnation at the mean, $j = 19, \ldots, 27$.

iii) normal (rapid) convergence to λ_1 in step with normal (rapid) convergence of $\theta_2^{(j)}$ to λ_2, $j = 28,\ldots,33$.

4. ERROR BOUNDS

The Kaniel-Paige-Saad error bounds, see (Parlett 1980, Chap. 12), give upper bounds on the "errors" $\theta_1^{(j)} - \lambda_1$, $k = 1,2,\ldots$ in terms of limited information about the other eigenvalues. Here we mention that each polynomial of degree j yields an upper bound on $\theta_1^{(j)} - \lambda_1$ (and $\lambda_n - \theta_j^{(j)}$) and, by specializing the choice, one can obtain bounds on $\theta_k^{(j)} - \lambda_k$ for higher values of k.

In (S. and V. 1987) a number of theorems are proved that yield expressions for the particular steps in a Lanczos run that separate the three phases exhibited in the numerical example. The analysis uses several nearby inner products and their associated Ritz values for comparison purposes. The natural comparison is the inner product in which λ_1 and λ_2 are replaced by

$$M := \tfrac{1}{2}(\lambda_1 + \lambda_2)$$

and the weight is adjusted according to $w_M = \sqrt{2}$. The associated Ritz values are denoted by $\bar{\theta}_i^{(j)}$, $i = 1,\ldots,j$.

We quote one result that determines for how many steps $\theta_1^{(i)}$ remains very close behind to $\bar{\theta}_1^{(i)}$.

Property 8.17.   Assume $0 < \varepsilon < \frac{1}{2}$ and

$$(\lambda_2 - \lambda_1) \sum_{k=3}^{i+1} \frac{1}{\lambda_k - \lambda_1} < \varepsilon \quad . \tag{8.18}$$

If, for $\beta \in (0, 1-\varepsilon)$,

$$\bar{\theta}_1^{(i)} - M \geq (\lambda_2 - \lambda_1)\varepsilon/2\beta \quad , \tag{8.20}$$

then

$$1 \geq \frac{\theta_1^{(i)} - M}{\bar{\theta}_1^{(i)} - M} \geq 1-\varepsilon-\beta \tag{8.21}$$

When ε and β have been chosen appropriately the theorem says that it is the eventual violation of (8.20) that breaks $\theta_1^{(i)}$'s bondage to $\bar{\theta}_1^{(i)}$. In the example $\varepsilon < .0131$ and so $\beta = 0.2$ is a valid choice that yields a value 0.33×10^{-5} for the right side of (8.20).

This is violated (barely) at Step 18 which Fig. 1 reveals as the end of Phase 1.

Different theorems are used to characterize the duration of the stagnation phase and the final convergence of $\theta_1^{(j)}$ and $\theta_2^{(j)}$ to λ_1 and λ_2.

5. THE POINT OF DISCOVERY

The many quantitative results in (S. and V. 1987) give a very accurate description of the behavior of $\theta_1^{(j)}$ and $\theta_2^{(j)}$, as j increases, in terms of the eigenvalues. However two distinct aspects of the process are fused in those theorems: the routine and the exceptional. We suggest an alternative description based on a single whole number, the step at which the comparison scheme must be discarded.

> There is an index j, *the point of discovery*, at which
> the process "perceives" that there must be a hidden
> eigenvalue in the vicinity of the stagnant Ritz values.
> Prior to step j the Ritz value development is routine —
> but for the comparison scheme wherein the close eigenvalues
> are pushed to their mean value. After step j the
> development is also routine — for the true regime.

In the Dutch example the point of discovery is step 24. However
this fact cannot be deduced from a local analysis of $\theta_1^{(j)}$ or even of
$\theta_1^{(j)}$, $\theta_2^{(j)}$, and $\theta_3^{(j)}$. Such a specific naming of the discovery point
is an artefact of our definition as well as of the situation itself.
Our choice, given in Section 8, has the attractive feature that no
knowledge of the true eigenvalues is needed to determine it. A simpler
definition of the discovery point, arising from notions to be given in
this section, would put the discovery at step 22 but this approach
requires knowledge of the eigenvalues to high accuracy.

We mention here that it is only *after* step 24 that $\theta_k^{(j)}$ approximates
λ_k for $k = 1,2,3,4$ and 5. Before that there is some mismatch between the
indices of Ritz values and of the eigenvalues they approximate. When we
say that the development is routine we mean that the Kaniel–Paige–Saad
error bounds are quite sufficient to describe how well certain eigen-
values are approximated at each (routine) step.

In order to justify this description of Ritz value behavior some
preparation is necessary.

Figure 2 shows the positions of the smaller Ritz values for the
comparison inner product (wherein λ_1 and λ_2 are moved to their mean).
The horizontal bars denote intervals guaranteed to contain an eigenvalue.
The configuration within the crowded interval [.2, 1.2], where 894 eigen-
values are located, is of little direct concern in this study and it is

not shown. The pattern revealed in Figure 2 should be considered routine
for all steps. The gaps between the six smallest eigenvalues are fairly
uniform.

Recall that $\{\chi_j\}_1^n$ denotes the sequence of orthogonal polynomials
associated with an inner product.

DEFINITION. For a given inner product and any $\xi \in \mathbb{R}$,

●   Left(j; ξ) := the number of zeros of χ_j that are less than ξ.

●   Right(j; ξ) := the number of zeros of χ_j that exceed ξ.

LEMMA 1. *For a fixed ξ both* Left(j;ξ) *and* Right(j;ξ) *are
monotone nondecreasing in j.*

This result is a direct consequence of the interlacing property.
In fact, for all j,

$$0 \leq \text{Left}(j{+}1;\ \xi) - \text{Left}(j;\ \xi) \leq 1 \ .$$

From now on we confine the discussion to discrete inner products with
n points of increase. What is of interest is the smallest index i at
which Left(i; ξ) attains its maximum value Left(n; ξ), the number of
eigenvalues less than ξ.

Figure 2 reveals that at step 18 the three smallest eigenvalues
are marked by the three smallest Ritz values while $\theta_4^{(18)} > \lambda_7$. This
suggests the following definitions:

     [Note that Left(k, $\frac{1}{2}(\lambda_i{+}\lambda_{i+1})$) $\leq$ i = Left(n, $\frac{1}{2}(\lambda_i{+}\lambda_{i+1})$)]

● $L(j)$:= min$\{i$: Left(j; $\frac{1}{2}(\lambda_i{+}\lambda_{i+1})$) $< i\}$,

● $R(j)$:= min$\{i$: Right(j, $\frac{1}{2}(\lambda_{n-i}{+}\lambda_{n-i+1})$) $< i\}$.

●   The *effective spread* at step $j := \lambda_{L(j)} - \lambda_{R(j)}$.

●   The eigenvalues outside $[\lambda_{L(j)}, \lambda_{R(j)}]$ are said to be

marked (at step j).

- The *effective condition number* for the assocaited Conjugate Gradient Algorithm (when $0 < \lambda_1 < \ldots < \lambda_n$) is

$$\lambda_{R(j)}/\lambda_{L(j)} \quad .$$

The following result gives substance to the words "marked" and "spread" used above.

LEMMA 2. *For each $k \geq j$ (but $k \leq n$),*

$$\theta_i^{(k)} \in [\lambda_i, \theta_i^{(j)}] \ , \quad \text{for } i = 1,2,\ldots,L(j)-1 \ ,$$

$$\theta_{k+1-i}^{(k)} \in [\theta_{j+1-i}^{(j)}, \lambda_{n+1-i}] \ , \quad \text{for } i = 1,2,\ldots,R(j)-1 \ ,$$

$$\theta_{\ell}^{(k)} \in [\lambda_{L(j)}, \lambda_{R(j)}] \ , \quad \text{for } L(j) \leq \ell \leq k+1-R(j) \ .$$

Proof. The first two assertions are direct consequences of the interlace property and the definition of $L(j)$ and $R(j)$. Recall that $\lambda_i = \theta_i^{(n)}$.

The third assertion is seen by contradiction. Suppose $\theta_{L(j)}^{(k)} < \lambda_{L(j)}$ for some $k \geq j$, then

$$\text{Left}(k, \lambda_{L(j)}) \geq L(j) > L(j)-1 = \text{Left}(n, \lambda_{L(j)}) \quad .$$

The first inequality follows from the definition of Left, the final equality follows from the definition of L.

This inequality contradicts the fact that $\text{Left}(j, \xi)$ is monotone nondecreasing in j (i.e. Lemma 1). Similarly $\theta_{k+1-R(j)}^{(k)}$ cannot exceed $\lambda_{R(j)}$ for any $k \geq j$.   □

Without knowledge of the $\{\lambda_i\}$ the indices $L(j)$ and $R(j)$ will also be unknown. We provide a computable alternative in Section 8. Here we seek only to explain Ritz value behavior. In order to quantify the late recognition of eigenvalue clusters we need one more elementary

consequence of the interlacing property. L(j) is the index of the

smallest unmarked eigenvalue. However it can happen that

Left(j, $\frac{1}{2}(\lambda_i+\lambda_{i+1})$) = i for some i > L(j). In fact, whenever the

largest index i for which Left(j, $\frac{1}{2}(\lambda_i+\lambda_{i+1})$) achieves its maximum

exceeds L(j), then the existence of hidden eigenvalues has been

discovered. This relatively uncommon event occurs in the Dutch example.

It can be seen at step 24 in Figure 3, which shows all the Ritz values

less than 0.2.

To formalize our explanation we make one more pair of definitions:

- $$\hat{L}(j) := \begin{cases} \max\{i: \text{Left}(j, \frac{1}{2}(\lambda_i + \lambda_{i+1})) = i\} \ , \\ 0 \ , \quad \text{if the set is empty} \end{cases}$$

- $$\hat{R}(j) := \begin{cases} \max\{i: \text{Right}(j, \frac{1}{2}(\lambda_{n+1-i} + \lambda_{n-i})) = i\} \ , \\ 0 \ , \quad \text{if the set is empty} \end{cases}$$

- The (theoretical) point of Discovery
$$:= \min\{j: \hat{L}(j) > L(j) \text{ or } \hat{R}(j) > R(j)\} \ .$$

We can call the configuration of Ritz values **routine** whenever
$$\hat{L}(j) + 1 = L(j) \ , \qquad \hat{R}(j) + 1 = R(j) \ .$$

6. DISCUSSION OF THE DUTCH EXAMPLE

Figure 3 reveals that, at step 22, $\theta_2 \approx \lambda_3$ and $\theta_3 \approx \lambda_4$. There is

every indication that this is a routine case. Comparison with Figure 2

(the comparison scheme) indicates that to within human vision the two

sets of Ritz values are identical through step 20. In the absence of

outside information one assumes that $\theta_3^{(k)}$ is going to converge to an

eigenvalue near 0.12. However a careful inspection of the computed Ritz

values show that $\theta_2^{(22)} = 0.0819858 < \lambda_3 = 0.082$. By Lemma 1,
Left(k,3) $\geq$ 2 for all k $\geq$ 22 and so there must be a hidden eigen-
value. So step 22 has a good claim to be the point of discovery.
Such a definition demands full knowledge of the configuration in which
case the existence of λ_2 will already be known. Our interest comes
from the practical situation where we want to make inferences about the
unknown eigenvalues from the known Ritz values.

At step 23 θ_4 slips noticeably below λ_3. However it is not
until step 24 that θ_4 is closer to λ_4 than to λ_3 and so θ_4 marks
λ_4 before θ_2 marks λ_2 at step 26. Table 1 reiterates the message of
Figure 3 but does so by means of natural numbers. Moreover, the profile
vector in the table shows the claim of step 22 to be the point of
discovery. Our use of midpoints between eigenvalues in the definition
of L and $\hat{L}$ accounts for the fact that $\hat{L}(j)$ does not exceed $L(j)$
until step 24. This value is corroborated in Section 8. The reward for
using midpoints is that, for $i < L(j)$, $\theta_i^{(j)}$ does actually mark λ_i.

The detailed results in (S. and V. 1987) do not reveal that at step
23 the close pair of eigenvalues near .034 caused the far off sequence
$\{\theta_3^{(k)}\}$ to be deflected from its apparent target $\lambda_4 = .127$ towards the
hidden pair and the concomittant deflection of $\{\theta_4^{(k)}\}$ to λ_4. This is
the essential insight provided by Figure 3 which also shows how mislead-
ing it can be to automatically associate $\theta_i^{(k)}$ with λ_i or $\theta_{k+1-i}^{(k)}$
with λ_{n+1-i}. There is no natural pairing of Ritz values at adjacent
steps. We offer our preferred way of pairing in Section 7.

Table 1
RITZ VALUE COUNTS

$$\text{Profile}(j) := [\text{Left}(j,\lambda_3), \ \text{Left}(j,\lambda_4), \ \text{Left}(j,\lambda_5), \ \text{Left}(j,\lambda_6)]$$

Step	$\hat{\text{L}}$	L	Profile			
10	0	1	1	1	1	2
11	0	1	1	1	2	2
12	0	1	1	2	2	2
15	0	1	1	2	2	3
16	0	1	1	2	3	3
21	1	2	1	2	3	4
22	1	2	2	2	3	4
23	1	2	2	3	4	4
→ 24	4	2	2	3	4	4
25	4	2	2	3	4	5
26	5	6	2	3	4	5
			.			
36	6	7	2	3	4	5

7. RITZ INTERVALS

To make a computable theory we shall replace Ritz values by Ritz intervals. This requires some extra notation. We write

$$T_j = \text{tridiag}\left[\alpha_1 \begin{array}{ccc} \beta_1 & \cdots & \beta_{j-1} \\ & \alpha_2 & \cdots & \alpha_j \\ \beta_1 & \cdots & \beta_{j-1} \end{array} \right] , \quad \begin{array}{l} \beta_i > 0 , \quad 1 < i < n, \\ \beta_n = 0 . \end{array}$$

The normalized eigenvectors are denoted by

$$T_j s_i^{(j)} = s_i^{(j)} \theta_i^{(j)} , \quad i = 1,\ldots,j , \quad \left\| s_i^{(j)} \right\|_2 = 1 .$$

The kth element of v is denoted by $v(k)$. Associated with each Ritz value $\theta_i^{(j)}$ is a Ritz intverval

$$I_i^{(j)} := \left[\theta_i^{(j)} - \beta_j |s_i^{(j)}(j)| , \ \theta_i^{(j)} + \beta_j |s_i^{(j)}(j)| \right] .$$

THEOREM (C. C. Paige). *For all* $k > j$ *(but* $k < n$) *there exists* ℓ *(depending on* k *and* i) *such that*

$$\theta_\ell^{(k)} \in I_i^{(j)} , \quad i = 1,\ldots,j .$$

This useful result is proved and discussed in (Parlett, 1980). The radius of $I_i^{(j)}$ can be given in terms of χ_j and χ_{j-1}:

$$\beta_{ji} := \beta_j |s_i^{(j)}(j)| = \left\{ \frac{\chi_{j-1}(\theta_i^{(j)})}{\|x_{j-1}\|^2} \Big/ \frac{\chi_j'(\theta_i^{(j)})}{\|x_j\|^2} \right\}^{\frac{1}{2}}$$

Here ϕ' denotes the derivative of ϕ. This interesting result does not seem to occur in (Szegö, 1939).

One consequence of this theorem is a useful pairing of the Ritz values at adjacent steps: Associate to $\theta_i^{(j)}$ the value $\theta_\ell^{(j+1)}$ that lies in $I_i^{(j)}$. If this rule does not specify ℓ (=i or i+1) uniquely then choose the $\theta_\ell^{(j+1)}$ that is closer to $\theta_i^{(j)}$.

DEFINITIONS.

- $I_i^{(j)}$ is *isolated* if $I_{i-1}^{(j)} \cap I_i^{(j)} = \emptyset = I_{i+1}^{(j)} \cap I_i^{(j)}$.

- $I_i^{(j)}$ is *marked* if either $I_i^{(j)}$ is isolated or $\theta_i^{(j)} \in I_\ell^{(j-1)}$ and $I_\ell^{(j-1)}$ is marked.

In the event that $I_\ell^{(j-1)}$ contains more than one j-level Ritz value we mark only the interval whose center is closer to $\theta_\ell^{(j-1)}$.

8. THE STANDING HYPOTHESIS

We use marked Ritz intervals to define analogues to $L(j)$ and $R(j)$. By convention $I_0^{(j)}$ and $I_{j+1}^{(j)}$ are marked.

DEFINTIONS.

- $\ell(j) = \max\{i: I_k^{(j)}$ is marked for all nonnegative $k < i\}$

- $r(j) = \max\{i: I_{j+1-k}^{(j)}$ is marked for all nonnegative $k < i\}$

- The *effective spread* (at step j) is $\bigcup_{i=\ell(j)}^{j+1-r(j)} I_i^{(j)}$

- The *standing hypothesis* (at step j) is that each marked Ritz interval contains exactly one eigenvalue.

Note that the effective spread may contain some interior isolated Ritz intervals. In contrast to $L(j)$ and $R(j)$, it is not the case that $\ell(j)$ and $r(j)$ are monotone nondecreasing in j. However we say that a configuration is *routine* whenever ℓ and r are monotone. In practice we use a violation of monotonicity to signal that the standing hypothesis may be false.

DEFINTION. The (computable) *point of discovery* =
$$\min\{j: \ell(j) < \ell(j-1) \text{ or } r(j) < r(j-1)\} .$$

Table 2 exhibits $l(j)$ for the Dutch example and the comparison inner product. The Ritz intervals are indicated by horizontal bars in Figures 2 and 3.

We now make some comments on the way in which monotonicity is violated. Consider step 23 in Figure 3. θ_3 and θ_4 both belong to $I_4^{(22)}$ but θ_3 is closer to the center. Hence $I_4^{(23)}$ is not marked and $l(23) = 4$. Again at step 24 both θ_3 and θ_4 belong to $I_3^{(23)}$, a marked interval. This time it is θ_4 that is closer to the center and so $I_3^{(24)}$ is unmarked while $I_4^{(24)}$ is marked. This causes $l(24)$ to drop down to 3 and signals a hidden eigenvalue. Similarly at step 25 it is $I_2^{(25)}$ that is unmarked and $l(25) = 2$. Not until step 36 is I_2 disjoint from I_1 and then $l(36)$ leaps up to 6.

The foregoing concepts find a use in comparing two discrete inner products whose internal structure is not known. We may say that two inner products *approximate each other* to degree j if

(1) the $l(i)$ and r(i) values first differ at $i = j+1$
and
(2) each marked Ritz interval contains the corresponding Ritz

value of the other scheme.

We can then say that the two inner products studied in this paper approximate each other for polynomials of degree less than 24.

9. THE $\|x_j\|$ CONNECTION

Figure 3 makes it clear that the actual step at which the presence of a hidden eigenvalue is revealed must depend on the actual distribution of a number of eigenvalues, not just those closest to the hidden one. One should expect a formula for this point to be complicated. Yet it turns out that the necessary global information is encoded in a computable

Table 2. MARKED RITZ INTERVAL COUNTS

Step	l (close eigenvalues)	l (comparison scheme)
7	2	2
13	3	3
18	4	4
23	4	4
→ 24	3	4
25	2	5
36	6	6

number; the radius of the relevant Ritz interval. That is one of the consequences of the theorem given below.

We abandon the Kaniel–Paige–Saad bounds and consider instead the relative importance of the contributions of a close pair of eigenvalues to $\|x_j\|^2$. The fluctuations of this contribution during the stagnation phase of the Dutch example are shown in Figures 5 and 6 by vertical spikes located at the eigenvalues. Each Δ indicates a Ritz value and a circle cuts off spikes that are too high. The sum of the heights of all spikes is 10 units.

THEOREM. *Let* $w_i = 1/n$, $i = 1,\ldots,n$, *and let*

- $h := (\lambda_{\ell+1} - \lambda_\ell)/2 \ll \min\{\lambda_\ell - \lambda_{\ell-1}, \ \lambda_{\ell+2} - \lambda_{\ell+1}\}$,

- $\theta := \theta_i^{(j)} = \frac{1}{2}(\lambda_\ell + \lambda_{\ell+1}) + \varepsilon h$, $-\frac{1}{2} < \varepsilon < \frac{1}{2}$, (this defines i),

- $C(\ell,j) := \left\{ \frac{1}{n} \chi_j^2(\lambda_\ell) + \frac{1}{n} \chi_j^2(\lambda_{\ell+1}) \right\}^{\frac{1}{2}}$

- $T_j s = s\theta$, $s^t s = 1$,

then

$$\frac{C(\ell,j)}{\|x_j\|} = \left\{ \frac{\sqrt{2/n}}{|s(1)|} \right\}\left(\frac{h}{\beta_j |s(j)|} \right) \left\{ 1 + \tfrac{1}{4}\varepsilon^2 + O(h\varepsilon) + O(h^2) \right\} .$$

Proof. The conditions on h permit the use of Taylor series to evaluate $C(\ell,j)$. Let $\chi_j^{(m)}$ denote $\chi_j^{(m)}(\theta)$, $m = 0,1,2$. Then

$$\chi_j(\lambda_\ell) = \chi_j + (\chi_j')(\lambda_\ell - \theta) + \tfrac{1}{2}(\chi_j'')(\lambda_\ell - \theta)^2 + O(h^3) ,$$

$$\chi_j(\lambda_{\ell+1}) = \chi_j + (\chi_j')(\lambda_{\ell+1} - \theta) + \tfrac{1}{2}(\chi_j'')(\lambda_{\ell+1} - \theta)^2 + O(h^3)$$

$$\chi_j^2(\lambda_\ell) = [\chi_j' h(1+\varepsilon)]^2 \left\{ 1 - \frac{\chi_j''}{\chi_j'} h(1+\varepsilon)\} + O(h^4) \right\} ,$$

$$\chi_j^2(\lambda_{\ell+1}) = [\chi_j' h(1-\varepsilon)]^2 \left\{ 1 + \frac{\chi_j''}{\chi_j'} h(1-\varepsilon)\} + O(h^4) \right. ,$$

$$C(\ell,j)^2 = \frac{2}{n} (h\chi_j')^2 \left\{ 1 + \varepsilon^2 - h\varepsilon \frac{\chi_j''}{\chi_j'} (3+\varepsilon^2)\} + O(h^4) \right. .$$

The following properties of tridiagonal matrices may be found in (Parlett 1980, p.129):

$$s(1)s(j)\chi_j' = \beta_1 \cdots \beta_{j-1} \quad ,$$

$$\|x_j\| = \beta_1 \cdots \beta_{j-1}\beta_j \quad .$$

Hence

$$\frac{C(l,j)}{\|x_j\|} = \sqrt{\frac{2}{n}} \left(\frac{h}{\beta_j s(1)s(j)} \right) \left\{ 1 + \varepsilon^2 - h\varepsilon(3 + \varepsilon^2) \frac{\chi_j''}{\chi'^2} \right\}^{\frac{1}{2}} + O(h^4) \quad .$$

The result follows. □

In Section 7 we mentioned that the ith Ritz interval of diameter $2\beta_j|s(j)|$ is guaranteed to contain an eigenvalue and thus

$$(1-\varepsilon)h < \beta_{ji} := \beta_j |s_i(j)| \quad .$$

It is the factor (h/β_{ji}) that controls $C(l,j)/\|x_j\|$. The constraint on ε in the defintion of h confines j to the stagnation phase in which the quantity $\frac{\sqrt{2/n}}{|s(1)|}$ will stay close to 1.

In the comparison scheme of the Dutch example (where $\lambda_1 = \lambda_2$), $|s(1)|$ converges rapidly to $\sqrt{2/n}$ as j increases whereas in the regime under study $|s(1)|$ does not decline to its limit $1/\sqrt{n}$ until $\theta_2^{(j)}$ marks λ_2 at step 27. At the beginning of stagnation the bound β_{ji} exceeds h by nearly an order of magnitude and $C(l,j)$ is insignificant. As the bound drops rapidly towards h the ratio $C(l,j)/\|x_j\|$ rises towards 1 and, *at that moment*, the minimizing property of the χ_i forces an extra Ritz value to be dispatched towards the hidden pair. However the consequent disruption of the distant Ritz values is reflected in a temporary *increase* in β_{ji} until the end of the stagnation phase. Thus $C(l,j)/\|x_j\|$ declines steadily only after the point of discovery. In the Dutch example $\beta_{28,1}/\beta_{23,1} > 10$.

> In other words, the discovery step is the
> one at which β_{ji} has a local minimum.

The preceding analysis shows that, in general, the discovery step increases only logarithmically with a decrease in h. If h is divided by 20 and all eigenvalues except for λ_ℓ and $\lambda_{\ell+1}$ remain fixed then the point of discovery is delayed by only 3 steps. Here is the reason.

This step may be characterized as the solution j to

$$\frac{C(\ell,j)}{\|x_j\|} = \kappa \quad (= 1/3 \text{ say}) \tag{A}$$

The proof of the preceding theorem shows that for small h,

$$\frac{C(\ell,j)}{\|x_j\|} = \sqrt{2/n} \; h \; \frac{|\chi_j'(\theta)|}{\|x_j\|} \quad . \tag{B}$$

Now,

$$\chi_j'(\theta) = \prod_{m \neq i}^{j} (\theta - \theta_m^{(j)}) \quad .$$

We approximate this expression by substituting marked eigenvalues for the p corresponding Ritz values and using a Chebyshev polynomial for the rest of the expression. Thus,

$$\chi_j'(\theta) \approx \prod_{k=1}^{P} (\theta - \lambda_k) \cdot T_{j-1-p}(\theta; [\lambda_{p+1}, \lambda_n]) \tag{C}$$

and the first factor is independent of j. For $\xi > 1$, $T_k(\xi) \approx \frac{1}{2}(\xi + \sqrt{\xi^2-1})^k$. Moreover, in most cases, when $j \ll n$,

$$\|x_j\| \approx \rho^j \tag{D}$$

for some fixed ρ. For the Dutch example, $\rho = 0.27$. In general, use (B),(C),(D) in (A) to see that the discovery value satisfies

$$h\kappa_1^j = \kappa_2$$

for complicated constants κ_1 and κ_2. Thus if h is divided by e^m then the discovery point will increase by $m/\ln \kappa_1$. This logarithmic dependence is illustrated in Figure 6 which shows the behavior on the Dutch example when the gap $\lambda_2 - \lambda_1$ is reduced to 1% of its original value.

Acknowledgments. I am grateful to my student, Mr. Yinsun Feng, who produced the figures which reveal the patterns so clearly.

REFERENCES

Kaniel, S. (1966). Estimates for some computational techniques in linear algebra. Mathematics of Computation 20, 369-378.

Paige, C. C. (1976). Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix. Journal of the Institute for Mathematics and Its Applications 18, 341-349.

Parlett, B. N. (1980). The symmetric eigenvalue problem. Prentice Hall, New Jersey.

Parlett, B. N., Simon, H. and Stringer, L. M. (1982). On estimating the largest eigenvalue with the Lanczos algorithm. Mathematics of Computation 38, 153-165.

Saad, Y. (1980). On the rate of convergence of the Lanczos and block Lanczos methods. SIAM Journal of Numerical Analysis 17, 687-706.

Szegö, G. (1939). Orthogonal polynomials. American Mathematical Society Colloquium Publication no. 23, 3rd edition, New York.

van der Sluis, A. and van der Vorst, H. (1987). The convergence behavior of Ritz values in the presence of close eigenvalues. Journal of Linear Algebra and Its Applications (Wilkinson issue), 88/89, 651-694.

FIGURE CAPTIONS

Figure 1.   Stagnation of a Ritz value at the mean .03405

Figure 2.   Ritz interval development; coincidence eigenvalues at
.03405 starting vector $(1\ 1\ \ldots\ 1\ 1)^t$

Figure 3.   Ritz interval development; close eigenvalues .0340
and .0341 starting vector $(1\ 1\ \ldots\ 1\ 1)^t$

Figure 4.   Relative magnitude of terms in the squared norm of the
j^{th} orthogonal polynomial n=900. Two close eigenvalues,
.0340 and .0341, next is .0820. Δ = Ritz value

Figure 5.   Relative magnitude of terms in the squared norm of the
j^{th} orthogonal polynomial n=900. Two close eigenvalues,
.0340 and .0341, next is .0820. Δ = Ritz value.

Figure 6.   Ritz interval development; eigenvalue gap = 10^{-6};
starting vector $(1\ 1\ \ldots\ 1\ 1)^t$

Figure 1.

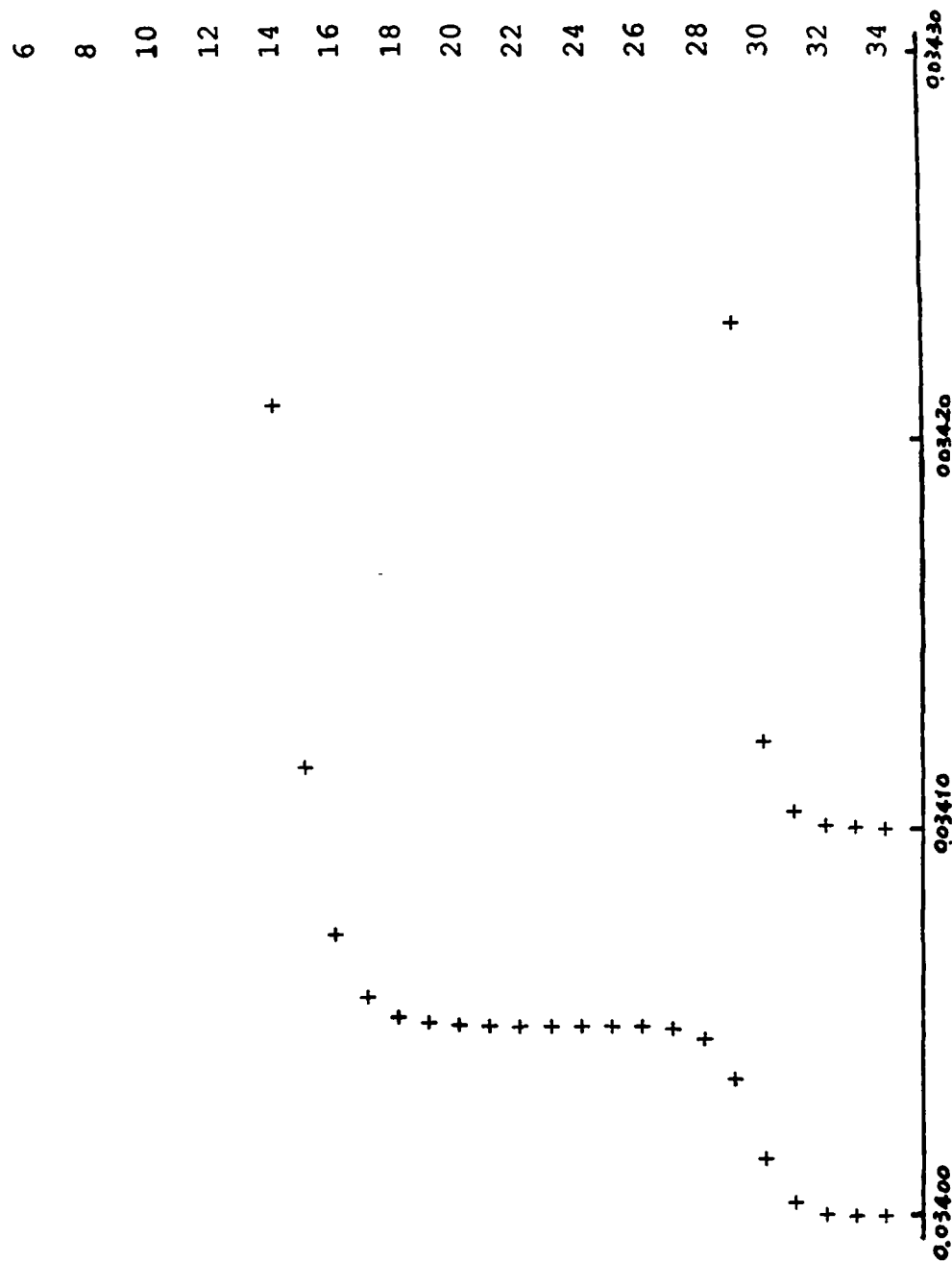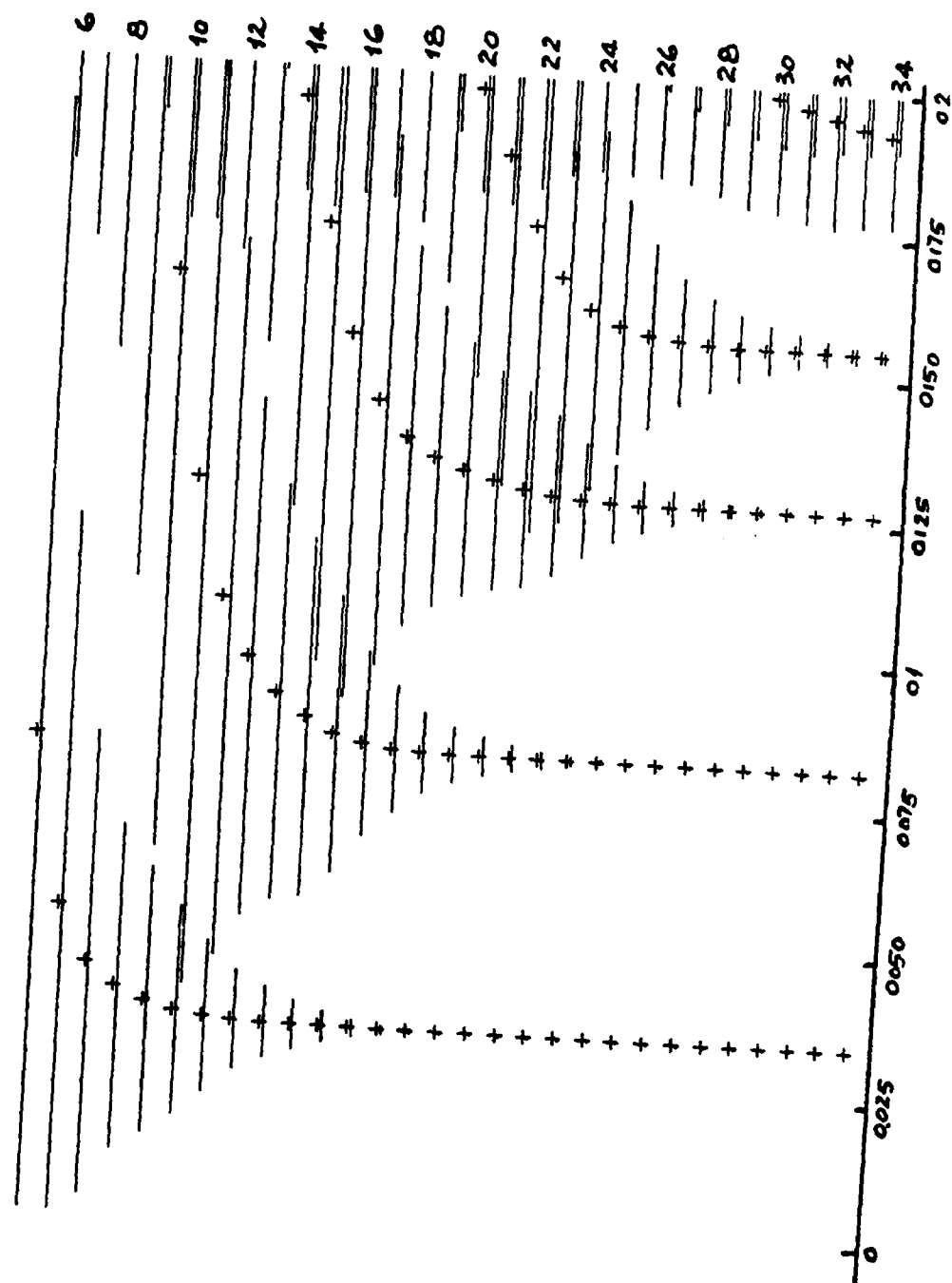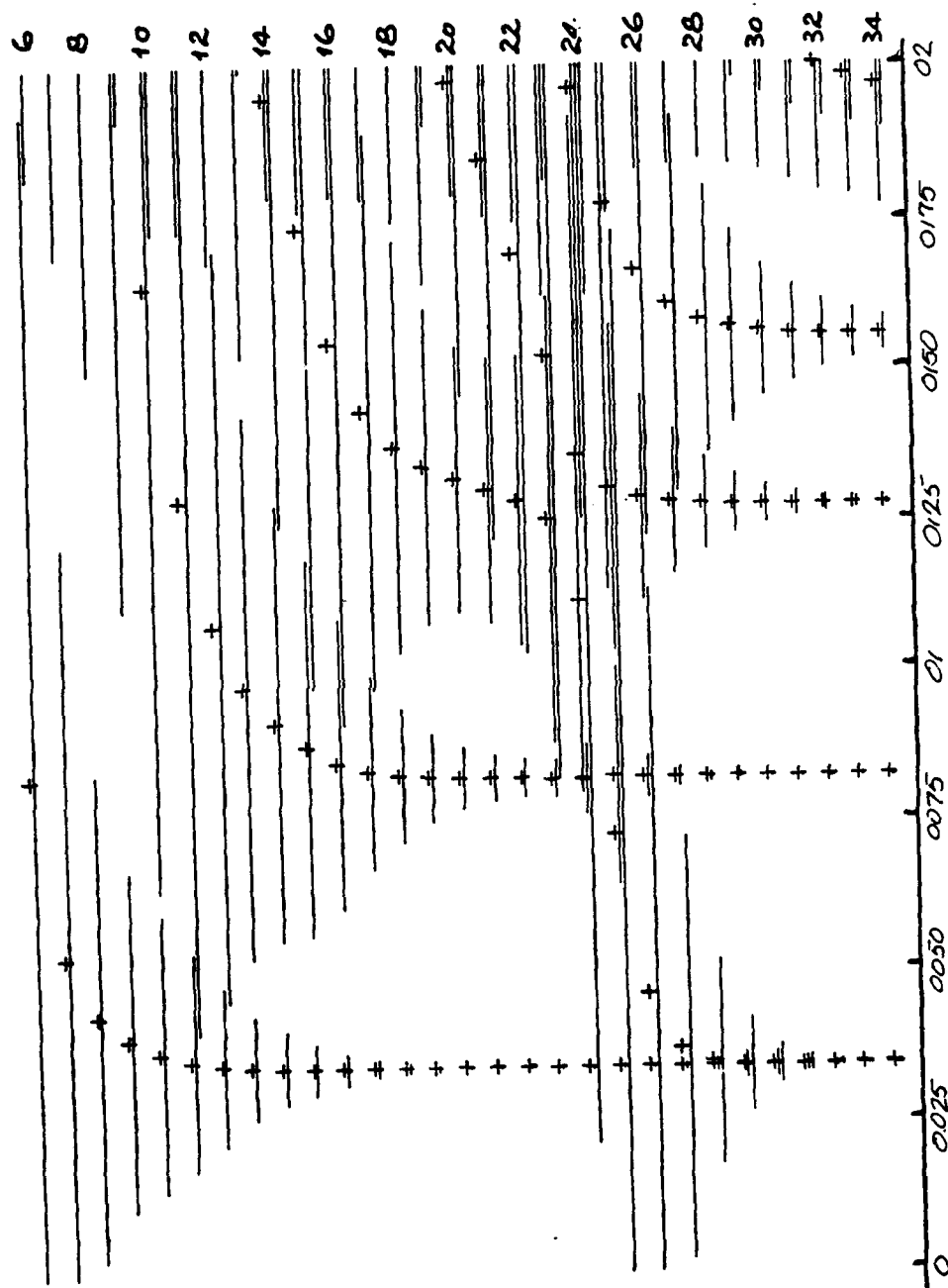## STAGNATION OF A RITZ VALUE
AT THE MEAN .03405

Figure 2.   RITZ INTERVAL DEVELOPMENT
COINCIDENT EIGENVALUES AT .03405
STARTING VECTOR (1 1 ... 1 1)*

Figure 3. **RITZ INTERVAL DEVELOPMENT**
CLOSE EIGENVALUES .0340 AND .0341
STARTING VECTOR $(1\,1\,...\,1\,1)^t$

25

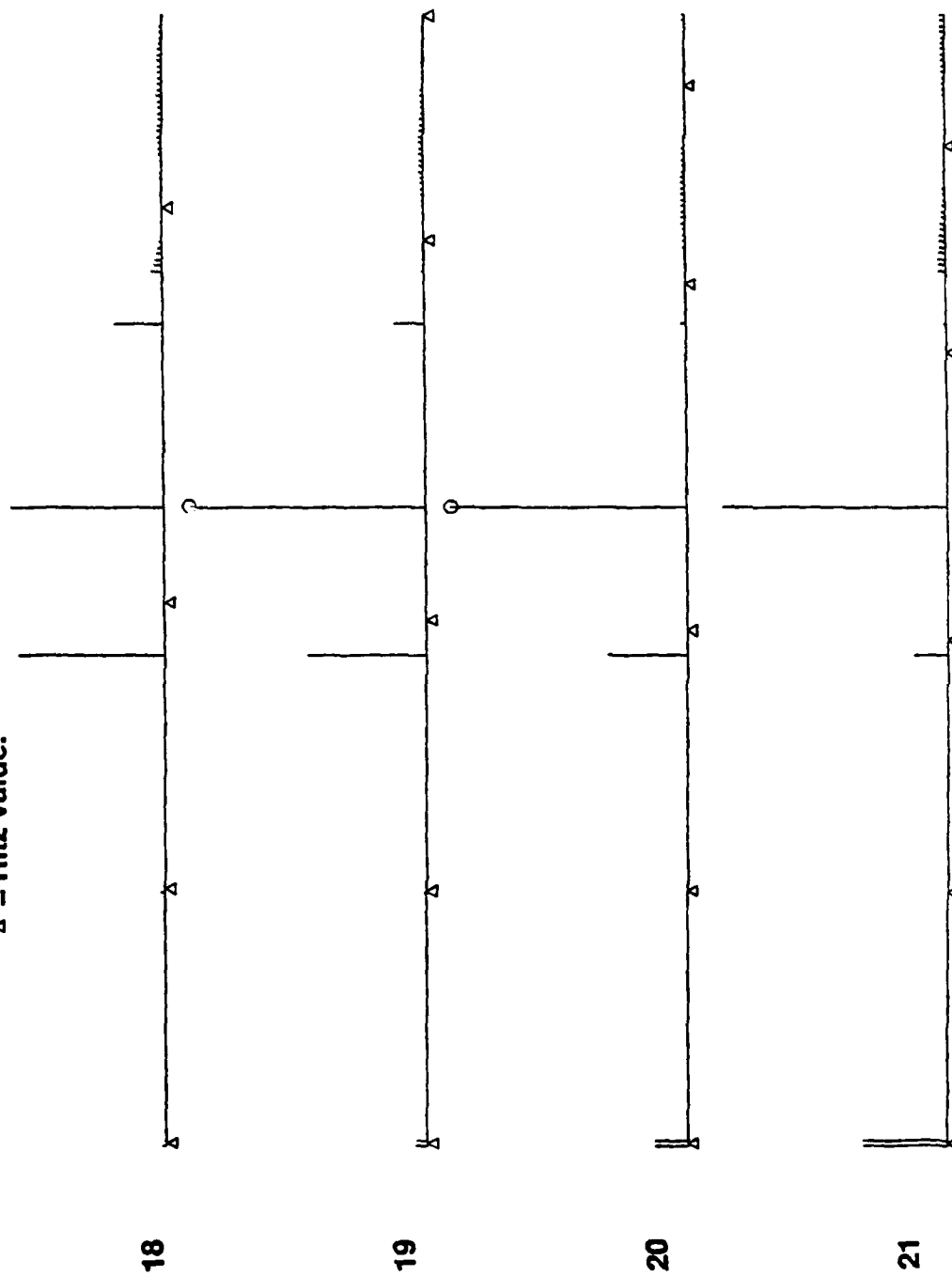Figure 4. Relative magnitude of terms in the squared norm of the j-th orthogonal polynomial n = 900. Two close eigenvalues, 0.0340 and 0.0341, next is 0.0820.
Δ = Ritz value.

Figure 5. Relative magnitude of terms in the squared norm of the j-th orthogonal polynomial n = 900. Two close eigenvalues, 0.0340 and 0.0341, next is 0.0820.
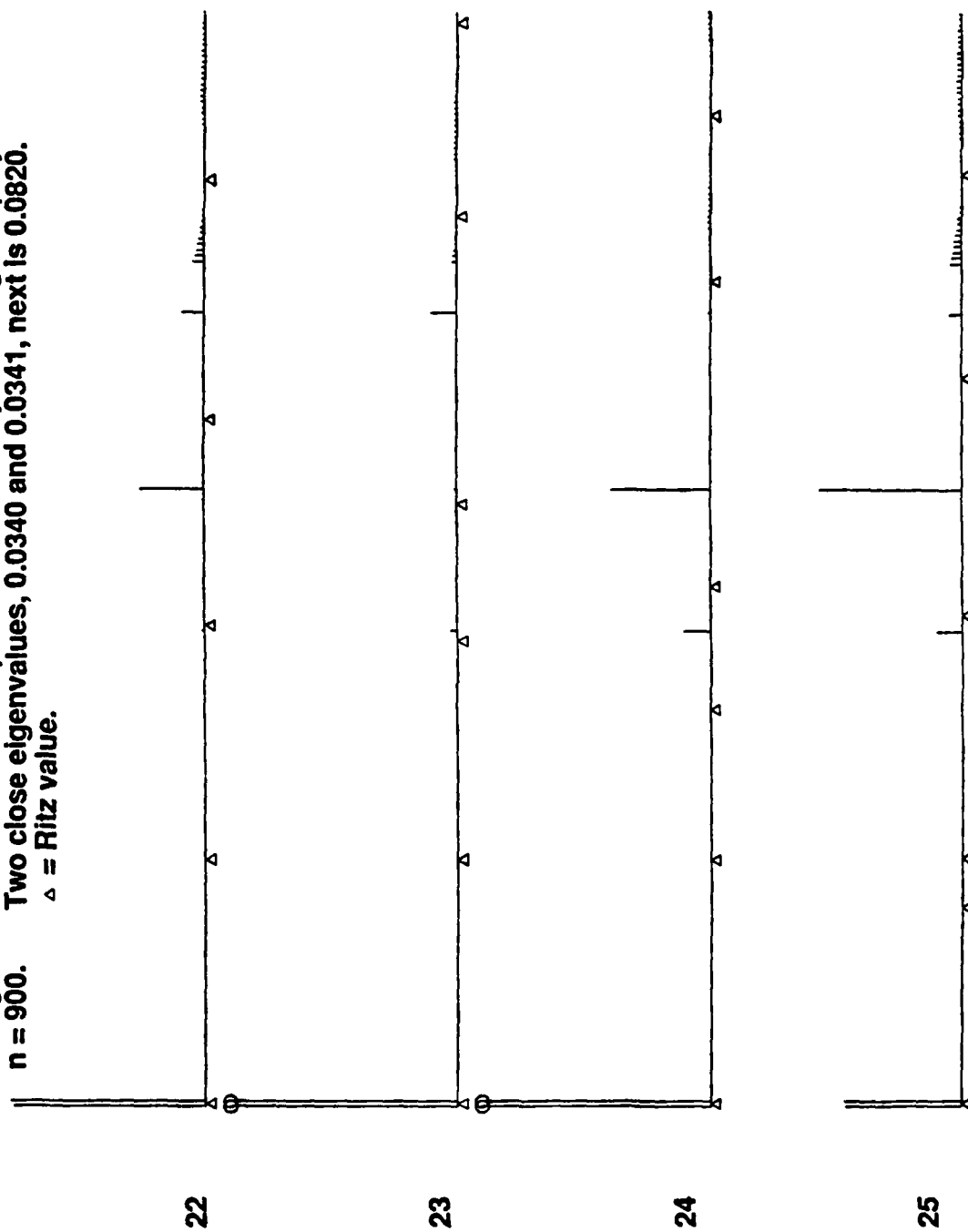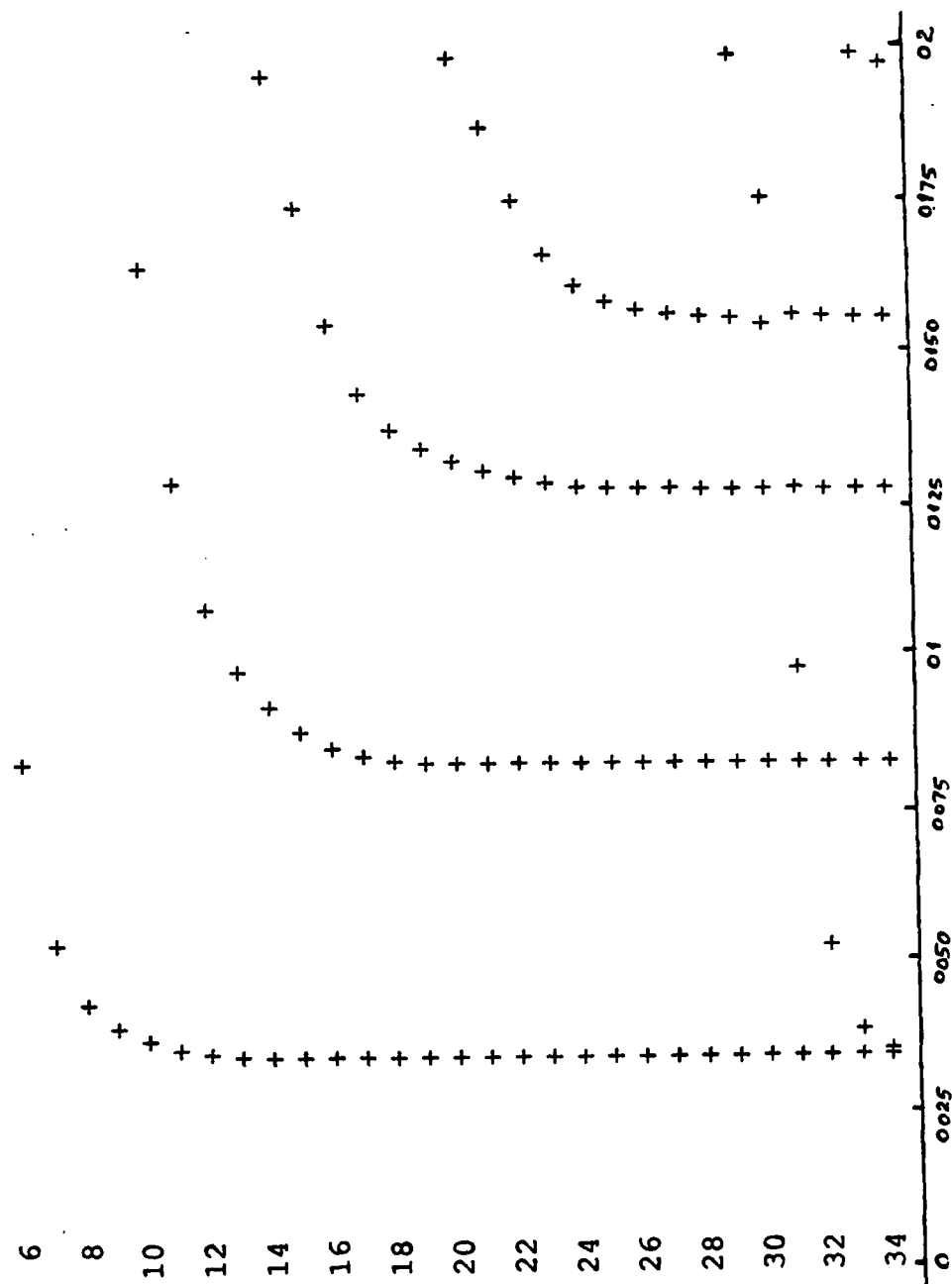△ = Ritz value.

Figure 6. RITZ INTERVAL DEVELOPMENT
EIGENVALUE GAP $= 10^{-6}$
STARTING VECTOR $(11 \dots 11)^t$

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
University of California, Berkeley	Unclassified
	2b. GROUP

3. REPORT TITLE

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Center for Pure & Applied Mathematics (CPAM) report, December 1987

5. AUTHOR(S) *(First name, middle initial, last name)*

Beresford N. Parlett

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1987	27	

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-85-K-0180	
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Mathematics Branch
	Office of Naval Research
	Washington, D.C. 20360

13. ABSTRACT

The Lanczos algorithm generates Ritz values in order to approximate eigenvalues. If some eigenvalues are clustered then a Ritz value may hover at a wrong value for a good number of steps. We study this phenomenon and focus on the point of discovery, the first step at which it is certain that there is a hidden eigenvalue in the vicinity of stabilized Ritz values. Both before and after this point the Ritz value behavior is routine — but for different eigenvalue configurations.

The "effective spread" at step j is an interval guaranteed to contain all unknown eigenvalues. The notion of "Ritz intervals" leads to a computable counterpart to the exact theory.

DD FORM 1473 (PAGE 1)
1 NOV 65

S/N 0101-807-6801